Nonlinear First-Order Partial Differential Equations by the Characteristic Method applied to The Hamilton-Jacobi Equation

Heikki von Koch

Master's thesis

Department of Mathematics and Statistics
University of Helsinki

Student number: 014449770

Abstract

This master's thesis has two main goals. First, to give a rigorous presentation of the method of characteristics without losing the important intuitive aspects that accompany it. Second, to use the method of characteristics to present the Hamilton-Jacobi equation, solve it in a reasonable manner and then introduce some of its most important properties. These topics will be accompanied by enlightening examples.

In the first part we develop the method of characteristics by transforming a first-order nonlinear PDE to a system of first-order ODE:s. We start by defining the notion of complete integrals as a sort of singleton solution to our PDE, and then combining these to a single family; an envelope encompassing the full solution. We then assume that the PDE itself can be written in a moving coordinate frame, and using this concept make an important assumption regarding the nature of the underlying curves at each moment. The full solution is finally achieved by weaving these curves to form the solution surface. To complete the theory, proper care needs to be taken of the boundary to make it compatible with our notion of curves. Lastly, all of the theory will be combined to make sure that the method actually produces well-defined local solutions.

With the method of characteristics developed, we have a look at the Hamilton-Jacobi equation. We will give sufficient conditions on when this initial-value problem can be solved with the Hopf-Lax formula. Based on this formula, a notion of weak solution will be given with its uniqueness proof. The Hamilton-Jacobi initial-value problem will be approached with the tools of variational calculus and convex analysis. These tools will be used to intimately link the Hamiltonian and Lagrangian by the means of the Legendre transform. The Hopf-Lax formula will then be constructed with the aim of solving the Hamilton-Jacobi initial-value problem. The formula is shown to have a useful functional identity as well as being Lipschitz continuous. Finally the uniqueness of the solution will be achieved by assuming semiconcavity from the initial function, or uniform convexity from the Hamiltonian.

The final chapter gives an insight as to how the developed theory can be further generalized and used. We will refer to some bibliography containing an abundance of further reading on semiconcave functions, optimal control theory and the Hamilton-Jacobi-Bellman equations.

Contents

Introduction

Partial differential equations (PDE:s) describe relations between an unknown function and its partial derivatives. The goal is to solve the unknown function u, where the full problem is stated in a form:

$$F(D^k u(x), D^{k-1}u(x), ..., Du(x), u(x), x) = 0 \quad x \in U,$$

with

$$F : \mathbb{R}^{n^k} \times \mathbb{R}^{n^{k-1}} \times \cdots \times \mathbb{R}^n \times \mathbb{R} \times U \to \mathbb{R}$$

given, and the parameter k standing for the order of the derivatives. The solution (if it exists) is found when we have discovered all u satisfying this equation, possibly only among those that meet certain boundary conditions on some part Γ of ∂U.

There are numerous of techniques for solving partial differential equations based on their nature. The method of characteristics is one of these techniques, targeting specifically first-order PDE:s. Its power lies in the fact that it turns the partial differential equation in to a system of ordinary differential equations. This is done by a careful analysis of the geometry of the PDE. This system is then open to a much more simpler, conventional analysis due to only depending on a single independent variable.

One important such a PDE is the Hamilton-Jacobi initial-value problem:

$$\begin{cases} u_t + H(Du) = 0 & \text{in } \mathbb{R}^n \times (0, \infty) \\ u = g & \text{on } \mathbb{R}^n \times \{t = 0\}. \end{cases}$$

The Hamilton-Jacobi equation can be thought to describe the time evolution of a mechanical system in the context of classical mechanics, but the ideas can be extended via the principle of least action to quantum physics as well. With the method of characteristics, this PDE can be broken down in to a dynamical system with a very specific structure: Hamilton's ODE. This special structure will allow the solution to be written in a closed form, the Hopf-Lax formula. With additional assumptions on the initial function g or on the Hamiltonian H, the Hopf-Lax formula produces a unique (weak) solution to this problem.

The purpose of this text is to characterise some of the solutions of first-order PDE:s, and solve them by using the method of characteristics. We will use this method to give an introduction to the Hamilton-Jacobi equations, and solve them with the help of calculus of variations and the Legendre transform. Finally we will see how to identify uniqueness of such a solution via appropriately defined weak solutions.

This text mostly follows Chapter 3 of Partial Differential Equations by Lawrence C. Evans [1].

1 Classification of solutions and the method of characteristics

The method of characteristics is a general technique for solving first-order differential equations by transforming them into systems of ordinary differential equations which are then more easily solved by conventional methods. Let us start by defining first-order PDE:s and making the notation clear.

Definition 1. *Let $x = (x_1, .., x_n) \in U \subseteq \mathbb{R}^n$ and $u : \bar{U} \to \mathbb{R}$, where U is open. A first order partial differential equation (PDE) for the unknown $u = u(x)$ is given by $F(Du, u, x) = 0$, where $F : \mathbb{R}^n \times \mathbb{R} \times \bar{U} \to \mathbb{R}$ is a given function and Du the the gradient of u.*

Depending on the nature of F, we call it linear, semilinear, quasilinear or fully nonlinear.

Definition 2. *The PDE $F(Du, u, x) = 0$ is called:*

1. *Linear if it has the form*

$$b(x) \cdot Du(x) + c(x)u(x) = f(x)$$

   for given functions b, c and f. This linear PDE is homogeneous if $f \equiv 0$.

2. *Semilinear if it has the form*

$$b(x) \cdot Du(x) + c(x, u(x)) = 0.$$

3. *Quasilinear if it has the form*

$$b(x, u(x)) \cdot Du(x) + c(x, u(x)) = 0.$$

4. *Fully nonlinear if it depends nonlinearly upon the highest order derivatives.*

It's customary to denote the gradient Du with p, and $u(x)$ with z so that the general form of the PDE reads

$$F = F(p, z, x) = F(p_1, ..., p_n, z, x_1, ..., x_n),$$

for $p \in \mathbb{R}^n$, $z \in \mathbb{R}$ and $x \in U$. We also henceforth assume that F is smooth and make the following conventions:

$$D_pF = (F_{p_1}, ..., F_{p_n}), \quad D_zF = F_z, \quad D_xF = (F_{x_1}, ..., F_{x_n}).$$

The PDE usually also comes with a boundary condition of the form

$$u = g \quad \text{on } \Gamma,$$

where $\Gamma \subseteq \partial U$ and $g : \Gamma \to \mathbb{R}$ are given. The problem is then referred to as a boundary value problem.

Fully nonlinear first-order partial differential equations are typically very hard to solve without placing some additional conditions on the PDE, and even then, in general, we are commonly only able to solve them locally (granted that we are able to solve them). The smoothness of the solution can usually be provided on a neighbourhood of the boundary but smoothness in the domain may or may not exist at all. This nonlinearity is also the reason why the problems do not often enjoy simple formulas for solutions.

1.1 Complete integrals and envelope solutions

Before we venture into the method of characteristics, we must first describe some simple solutions and then rigorously assert that we can construct new, more complex solutions from them (which is what the method of characteristics is ultimately based on).

Let $A \subset \mathbb{R}^n$ be an open set. Assume that for each parameter $a = (a_1, .., a_n) \in A$ we have a C^2 solution $u = u(x; a)$ of the PDE:

$$F(Du, u, x) = 0. \tag{1.1.1}$$

Definition 3. *A C^2 function $u = u(x; a)$ is called a complete integral in $U \times A$ provided*

1. *$u(x; a)$ solves the PDE for each $a \in A$*

and

2. *rank$(D_a u, D^2_{xa} u) = n$, where $x \in U$, $a \in A$.*

Remark. *We use the following shorthand notation in condition 2.*

$$(D_a u, D^2_{xa} u) := \begin{pmatrix} u_{a_1} & u_{x_1 a_1} & u_{x_2 a_1} & \dots & u_{x_n a_1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ u_{a_n} & u_{x_1 a_n} & u_{x_2 a_n} & \dots & u_{x_n a_n} \end{pmatrix}_{n \times (n+1)}.$$

Thus a complete integral is a solution of a PDE of the first order that contains as many arbitrary constants as there are independent variables. The second condition guarantees that the solution $u(x; a)$ depends on all of the

n independent parameters a (this can be seen by assuming otherwise and then checking that the determinant of each submatrix constructed from the assumed new solution equals zero; consequently the rank is strictly less than n).

Example 1. *The very important Hamilton-Jacobi equation that forms an alternative formulation of classical mechanics in its simplest form is the partial differential equation*

$$u_t + H(Du) = 0,$$

where $H : \mathbb{R}^n \to \mathbb{R}$ *is a given Hamiltonian and* $u = u(x)$. *By setting* $x = (x_1, .., x_n) \in \mathbb{R}^n$, $t \in \mathbb{R}$, $t = x_{n+1}$ *and* $Du = D_x u = (u_{x_1}, ..., u_{x_n})$; *a complete integral of the PDE is*

$$u(x, t; a, b) = a \cdot x - tH(a) + b,$$

where $a \in \mathbb{R}^n$, $b \in \mathbb{R}$ *and* $t \geq 0$. *This can be verified by simple calculations:* $u_t = -H(a)$, $Du = a$, $D_a u = x$, $D^2_{xa} = diag(1_1, ..., 1_n)$, *and so* $rank(D_a u, D^2_{xa} u) = n$.

Remark. *The above construction can also be acquired by assuming that the variables of* u *can be separated additively. Let us look for a solution of the form:*

$$u(x, t) = w(x) + v(t) \quad x \in \mathbb{R}^n, t \geq 0.$$

Now

$$0 = u_t(x, t) + H(Du(x, t)) = v'(t) + H(Dw(x))$$

if and only if

$$H(Dw(x)) = \alpha = -v'(t) \quad t > 0$$

for some constant α. *Consequently if* $H(Dw) = \alpha$ *for some* $\alpha \in \mathbb{R}$, *then*

$$u(x, t) = w(x) - \alpha t + b$$

will solve

$$u_t + H(Du) = 0$$

for any constant b. *Setting* $w(x) = a \cdot x$ *for some* $a \in \mathbb{R}^n$ *and* $\alpha = H(a)$, *we get the same solution as above.*

Very often the convention $t = x_{n+1}$ is used in literature to give a special meaning to the last position. This is because the parameter is usually reserved for time, which is one of the hugely important single variables used in most PDE:s. This text will follow this convention.

We shall next construct new solutions from complete integrals. These solutions will turn out to be somewhat more complicated due to the fact that they will depend on an arbitrary function of $n-1$ variables instead of just n parameters as in the definition of the complete integral. In general, the new solutions will come out as envelopes of complete integrals of other m-parameter families of solutions.

Definition 4. *Let $u = u(x;a)$ be a C^1 function of $x \in U$, $a \in A$, where $U \subset \mathbb{R}^n$ and $A \subset \mathbb{R}^m$ are open sets. Consider the equation*

$$D_a u(x;a) = 0. \tag{1.1.2}$$

Suppose that we can solve (1.1.2) for the parameter a as a C^1 function of x, $a = \phi(x)$ and so:

$$D_a u(x;\phi(x)) = 0. \tag{1.1.3}$$

We then call

$$v(x) := u(x;\phi(x)) \tag{1.1.4}$$

the envelope of the functions $\{u(\cdot;a)\}_{a\in A}$.

It's worth noting that the function u is now assumed to be C^1 instead of C^2, which is much less restrictive. The assumption that the parameters a are a function of x also reduces the amount of independent variables from $n+m$ to just $n+1$ (since the function ϕ itself varies).

With the help of the newly defined envelopes we can now construct new solutions of the nonlinear first-order PDE.

Theorem 1. *Suppose for each $a \in A$ as above that $u = u(\cdot;a)$ solves the partial differential equation (1.1.1). Assume further that the envelope v (sometimes called a singular integral of (1.1.1)), defined by (1.1.3) and (1.1.4) above, exists and is a C^1 function. Then v solves (1.1.1) as well.*

Proof. Since $v(x) = u(x;\phi(x))$, we have for $i = 1, ..., n$:

$$v_{x_i}(x) = u_{x_i}(x;\phi(x)) + \sum_{j=1}^{n} u_{a_j}(x,\phi(x))\phi^j_{x_i}(x) = u_{x_i}(x;\phi(x)),$$

according to (1.1.3). Thus for each $x \in U$,

$$F(Dv(x), v(x), x) = F(Du(x;\phi(x)), u(x;\phi(x)), x) = 0.$$

□

The concept of envelopes become more clear when thought geometrically. For each $x \in U$, the graph of $v(x)$ is everywhere tangent to the graph of some member of the family $u(x;a)$ for $a = \phi(x)$, and so $Dv = D_x u(\cdot; a)$ at x. A point on the envelope can be thus thought as being the intersection of two infinitesimally adjacent curves. Therefore it's clear that the individual members of the envelope family need to be continuously differentiable, and these points of tangency together form the whole envelope. A good physical example of an envelope comes from geometrical optics in the form of caustics; the envelope of a family of light rays.

By varying the above construction, we can generate even more possible solutions from a complete integral. This is done by restricting the parameter a to have a form $a = (a', h(a'))$ for a suitable function h depending on $n-1$ variables; the solution then depends on $n-1$ variables instead of the arbitrary n ones (and an explicit function h).

Definition 5. *Choose any open set $A' \subset \mathbb{R}^{n-1}$ and any C^1 function $h : A' \to \mathbb{R}$, so that the graph of h lies within A. Set: $a = (a_1, ..., a_n) = (a', a_n)$ for $a' = (a_1, ..., a_{n-1}) \in \mathbb{R}^{n-1}$. The general integral (depending on h) is the envelope $v' = v'(x)$ of the functions*

$$u'(x; a') = u(x; a', h(a'))$$

provided this envelope exists and is C^1.

Example 2. *Let us return to the Hamilton-Jacobi equation of the previous example, and set $H(p) = |p|^2$ and $h \equiv 0$ (so now $b = 0$). We have that*

$$u'(x, t; a) = a \cdot x - t|a|^2,$$

and so we can calculate the envelope using condition (1.1.3). That is $D_a u' = x - 2ta = 0$, from which we get $a = \frac{x}{2t}$; and so

$$u'(x,t) = x \cdot \frac{x}{2t} - t\left|\frac{x}{2t}\right|^2 = \frac{|x|^2}{4t}, \quad x \in \mathbb{R}^n,\ t > 0$$

solves the Hamilton-Jacobi equation $u_t + |Du|^2 = 0$.

Caution must always be exercised when dealing with solutions of a PDE. One could easily (erroneously) believe to have found all the solutions with the help of the function h defined above. For example think of a PDE that can be composed to two separate PDE:s:

$$F(Du, u, x) = F_1(Du, u, x)F_2(Du, u, x) = 0.$$

Now having a complete integral of the PDE $F_1(Du, u, x) = 0$ and finding a general integral corresponding to any function h, we will have missed all the solutions of the PDE $F_2(Du, u, x) = 0$, and hence solutions of the whole PDE $F(Du, u, x) = 0$.

1.2 Derivation of the characteristic equations

Let us consider a general first-order PDE

$$F(Du, u, x) = 0 \quad \text{in } U \tag{1.2.1}$$

along with a boundary condition

$$u = g \quad \text{on } \Gamma, \tag{1.2.2}$$

where $\Gamma \subseteq \partial U$ and $g : \Gamma \to \mathbb{R}$ are given, and F and g are smooth. The intuition behind the method of characteristics is to change the variables so that equation (1.2.1) can be written in a moving coordinate frame. This leads to the concept of curves, which can be seen as propagators of information that the PDE holds. The method reduces the PDE to a family of ODE:s from which the solution can be obtained by integrating from the given initial data on a hypersurface. The efficiency of the method lies in the fact that we will be able to solve a first-order nonlinear PDE by reducing it to first-order ODE:s with the initial data inherited from the PDE. These ODE:s will in general be easier to solve and analyse than the original PDE.

Let $u \in C^2$ solve (1.2.1), (1.2.2), and fix any point $x \in U$. Assume that the curve can be parametrized: $x(s) = (x^1(s), ..., x^n(s))$, where s lies in some interval $I \subseteq \mathbb{R}$. We shorten and simplify the notation by setting:

$$z(s) \coloneqq u(x(s)), \quad p(s) \coloneqq Du(x(s)),$$

where $p(s) = (p^1(s), ..., p^n(s))$ and $p^i(s) = u_{x_i}(x(s))$, $i = 1, .., n$. Now we need to choose the curve in such a way that we can compute z and p. Differentiating the original PDE with respect to x_i and the expression for the gradient with respect to s we obtain:

$$\sum_{j=1}^{n} F_{p_j}(Du, u, x)u_{x_j x_i} + F_z(Du, u, x)u_{x_i} + F_{x_i}(Du, u, x) = 0$$

and

$$\dot{p}^i(s) = \sum_{j=1}^{n} u_{x_i x_j}(x(s))\dot{x}^j(s).$$

Now we see that in order for us to get rid of the pesky second derivatives of u, we need an additional assumption; let us set:

$$\dot{x}^j(s) = F_{p_j}(p(s), z(s), x(s)), \quad j = 1, .., n.$$

Since this assumption is based on the evaluation at $x = x(s)$, let us also evaluate the differentiated PDE at it. That is:

$$\sum_{j=1}^{n} F_{p_j}(p(s), z(s), x(s))u_{x_j x_i} + F_z(p(s), z(s), x(s))p^i(s) + F_{x_i}(p(s), z(s), x(s)) = 0.$$

Now combining we get:

$$\begin{aligned}\dot{p}^i(s) &= \sum_{j=1}^{n} u_{x_i x_j}(x(s))\dot{x}^j(s) = \sum_{j=1}^{n} u_{x_i x_j}(x(s))F_{p_j}(p(s), z(s), x(s)) \\ &= -F_{x_i}(p(s), z(s), x(s)) - F_z(p(s), z(s), x(s))p^i(s).\end{aligned}$$

Finally we differentiate $z(s)$ with respect to s:

$$\dot{z}(s) = \sum_{j=1}^{n} u_{x_j}(x(s))\dot{x}^j(s) = \sum_{j=1}^{n} p^j(s)F_{p_j}(p(s), z(s), x(s)).$$

Gathering the obtained equations into a system, we have the *characteristic equations* of the nonlinear first-order PDE:

$$\begin{cases} \text{(a) } \dot{p}(s) = -D_x F(p(s), z(s), x(s)) - D_z F(p(s), z(s), x(s)) \cdot p(s) \\ \text{(b) } \dot{z}(s) = D_p F(p(s), z(s), x(s)) \cdot p(s) \\ \text{(c) } \dot{x}(s) = D_p F(p(s), z(s), x(s)). \end{cases} \tag{1.2.3}$$

This is a system of $2n+1$ first-order ODE:s for which $F(p(s), z(s), x(s)) = 0$, $s \in I$. The function F is solved by the parameters z, p and x where z gives the values of the original solution u along the curve, p records the gradient and x is itself the curve. The curve $x(s)$ is the projection of the full characteristic $(p(s), z(s), x(s)) \subset \mathbb{R}^{2n+1}$ onto the "physical" region $U \subset \mathbb{R}^n$, which is why it will be referred to as the projected characteristic. We can collect all the aforementioned data to a theorem:

Theorem 2. *Let $u \in C^2(U)$ solve the nonlinear first-order partial differential equation (1.2.1) in U. Assume $x(\cdot)$ solves the ODE (1.2.3)(c), where $p(\cdot) = Du(x(\cdot))$ and $z(\cdot) = u(x(\cdot))$. Then $p(\cdot)$ solves the ODE (1.2.3)(a) and $z(\cdot)$ solves the ODE (1.2.3)(b), for those $s \in I \subseteq \mathbb{R}$ such that $x(s) \in U$.*

Notice how the system (1.2.3) is closed; the equations fully describe the PDE with just first-order ODE:s. The higher order derivatives of u could be eliminated by setting $\dot{x} = D_p F$. This assumption in a way encompasses the meaning of the characteristic curves and their relation to the solution u.

Now that we have found a family of curves lying in U, we want to connect each of them with points (preferably separate) on the boundary Γ. With the boundary condition g given, we can calculate u at the boundary; and then by connecting the projected characteristic $x(s)$ with a point x^0 on Γ, we get the curves along which we fully compute u from. For this we need to make sure that our boundary is well behaved, and then make it compatible with our system of characteristic equations. This will ensure that u can be at least locally determined in regions of $\bar{U}$.

1.3 Boundary conditions

Boundary conditions tell us where the characteristic curves "start" from. Suppose that a function with constant values propagating along a characteristic curve passes through two different points on the boundary; we must ensure that the boundary data matches with this. That is if $x(s)$ is a projected characteristic and $x^0 \neq x(s_0)$ with both $x^0, x(s_0) \in \Gamma$, then we must make sure that $g(x^0) = g(x(s_0)$. Since our aim is to recover information about u by solving the characteristic ODE, we need to examine the boundary condition closer. First we show that we can assume that boundary is flat near x^0 by straightening it if necessary. This will somewhat simplify the notation of the conditions.

1.3.1 Straightening the boundary

Fix $x^0 \in \Gamma$ and find smooth mappings (see appendix C.1 from [1]) $\Phi, \Psi : \mathbb{R}^n \to \mathbb{R}$ with $\Psi = \Phi^{-1}$ such that Φ "straightens out" ∂U near x^0. We need a C^1 solution $u : U \to \mathbb{R}$, so let us write $V \coloneqq \Phi(U)$ and set

$$v(y) \coloneqq u(\Psi(y)) \quad y \in V,$$

so that

$$u(x) = v(\Phi(x)) \quad x \in U.$$

We illustrate that the PDE, function G, associated with v is a PDE of the "same form" as the PDE for u. Calculating the derivatives of u shows this;

$$u_{x_i} = \sum_{k=1}^{n} v_{y_k}(\Phi(x))\Phi^k_{x_i}(x) \quad i = 1, .., n$$

which equates in vector form to

$$Du(x) = Dv(y)D\Phi(x).$$

Plugging this into (1.2.1) implies that

$$0 = F(Du(x), u(x), x) = F(Dv(y)D\Phi(\Psi(y)), v(y), \Psi(y)).$$

Since $\Phi\Psi = \Phi\Phi^{-1} = id$, this has a simple form

$$G(Dv(y), v(y), y) = 0 \quad \text{in } V.$$

Defining $v(y) = h(y) \coloneqq g(\Psi(y))$ on $\Delta \coloneqq \Phi(\Gamma)$, we therefore have the transformed ("straightened" boundary) problem

$$\begin{cases} G(Dv(y), v(y), y) = 0 & \text{in } V \\ v = h & \text{on } \Delta. \end{cases} \tag{1.3.1}$$

1.3.2 Boundary compatibility conditions

As shown in the previous section, for a a given fixed $x^0 \in \Gamma$, we can assume that Γ is flat near it (lies in the plane $\{x_n = 0\}$). We now need to define our initial conditions so that the ODE truly produces solutions along the characteristics; at least near x^0. The point x^0 is given and we are to determine the proper initial conditions

$$p(0) = p^0,\ z(0) = z^0,\ x(0) = x^0,$$

such that they obey the ODE. Since $z(s) = u(x(s))$, we should have that

$$z^0 = z(0) = u(x(0)) = u(x^0) = g(x^0). \tag{1.3.2}$$

The last thing we need to deal with is p^0. The straightened boundary condition implies that $u(x_1, .., x_{n-1}, 0) = g(x_1, ..., x_{n-1})$ near x^0 and so we differentiate it to find that

$$u_{x_i}(x^0) = g_{x_i}(x^0) \quad i = 1, ..., n-1.$$

Finally since the original PDE needs to hold true, we have our compatibility conditions on p^0:

$$\begin{cases} p_i^0 = g_{x_i}(x^0) \quad i = 1, ..., n-1 \\ F(p^0, z^0, x^0) = 0. \end{cases} \tag{1.3.3}$$

This is a system of n equations with n unknowns. The combination of (1.3.2) and (1.3.3) are called the *compatibility conditions*. A triple $(p^0, z^0, x^0) \in \mathbb{R}^{2n+1}$ verifying said conditions is called *admissible*. It's noteworthy that z^0 is uniquely determined by the boundary condition and by the point x^0, but it may happen that the vector p^0 may not either be unique or even exist at all (this is determined by the equations (1.3.3)).

1.3.3 Noncharacteristic boundary data

This section provides the condition on when we will be able to locally solve a first-order PDE near the boundary using the method of characteristics. So far we have the boundary in control for a single argument, but what we really need is a solution of the characteristic ODE for the nearby points as well. Assume that $x^0 \in \Gamma$ where Γ is straightened, and that the triple (p^0, z^0, x^0) is admissible. We need to perturb the point (p^0, z^0, x^0) in such a manner that will keep the compatibility conditions in check.

Fix $y = (y_1, ..., y_{n-1}, 0) \in \Gamma$ near x^0 and solve the characteristic ODE (1.2.3) subject to the initial conditions

$$p(0) = q(y),\ z(0) = g(y),\ x(0) = y. \tag{1.3.4}$$

It's clear that our function $q(\cdot) = (q^1(\cdot), ..., q^n(\cdot))$ has to also satisfy the compatibility conditions with $q(x^0) = p^0$; that is for all $y \in \Gamma$ close to x^0 we must have

$$q(x^0) = p^0, \quad \begin{cases} q^i(y) = g_{x_i}(y) \quad i = 1, ..., n-1 \\ F(q(y), g(y), y) = 0. \end{cases} \tag{1.3.5}$$

What can we say about the solvability of such a function q restricted with the conditions of (1.3.5)?

Lemma 1. *(Noncharacteristic boundary conditions). There exists a unique solution $q(\cdot)$ of (1.3.5) for all $y \in \Gamma$ sufficiently close to x^0, provided*

$$F_{p_n}(p^0, z^0, x^0) \neq 0. \tag{1.3.6}$$

The admissible triple (p^0, z^0, x^0) is said to be *noncharacteristic* if (1.3.6) holds. This shall be henceforth assumed.

Proof. Let $y = (y_1, ..., y_n) \in \mathbb{R}^n$ and apply the Implicit Function Theorem to the mapping

$$G : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^n, \quad G(p, y) = (G^1(p, y), ..., G^n(p, y)),$$

where

$$\begin{cases} G^i(p, y) = p_i - g_{x_i}(y), \quad \text{for} \quad i = 1, .., n-1 \\ G^n(p, y) = F(p, g(y), y). \end{cases}$$

Now according to (1.3.2) and (1.3.3), we have $G(p^0, x^0) = 0$. By simple calculations

$$D_p G(p^0, x^0) = \begin{pmatrix} 1 & \dots & 0 & 0 \\ \vdots & \ddots & \vdots & \vdots \\ 0 & \dots & 1 & 0 \\ F_{p_1}(p^0, z^0, x^0) & \dots & \dots & F_{p_n}(p^0, z^0, x^0) \end{pmatrix}_{n \times n}$$

and $\det D_p G(p^0, x^0) = F_{p_n}(p^0, z^0, x^0) \neq 0$, as per the characteristic condition (1.3.6). The Implicit Function Theorem thus ensures we can uniquely solve the identity $G(p, y) = 0$ for $p = q(y)$, provided y is close enough to x^0. □

Remark. *If the boundary Γ has not been flattened near x^0, the generalized condition that Γ be noncharacteristic reads*

$$D_p F(p^0, z^0, x^0) \cdot v(x_0) \neq 0,$$

$v(x_0)$ denoting the outward unit normal to ∂U at x^0.

The noncharacteristic condition (1.3.6) (or its general vector version) has a simple meaning that's apparent from the proof of Lemma 1. The existence of unique characteristic curves as a solution near the boundary requires the invertibility of the Jacobian of G to ensure the uniqueness of the function q defining the boundary neighbourhood. This condition on the boundary forces you to work inside the defined fixed domain; if the condition forced you out of the domain, it would then by definition not be a boundary condition anymore but a part of the solution. It follows that for any given PDE there might exist initial curves for which the Jacobian vanishes (Lemma 1 does not hold), since the PDE and the initial conditions do not depend on each other. The noncharacteristic condition "pushes" the solutions out of the $\{x_n = 0\}$ plane.

1.4 Local existence of solutions

There is no reason to believe that our PDE could even be locally solvable. A simple example of such a system can be constructed using some mixed derivatives with specific initial conditions. For example the system:

$$u_x = yu, \quad u_y = 0 \tag{1.4.1}$$

$$u^0 = u(x_0, y_0) = 1, \quad u_x^0 = u_x(x_0, y_0) = y_0, \quad u_y^0 = u_y(x_0, y_0) = 0, \tag{1.4.2}$$

doesn't have local solutions $u \in C^2$ (the only solution being the trivial one $u(x, y) \equiv 0$) as can be seen from:

$$u_{xy} = (yu)_y = yu_y + u = 0.$$

The initial conditions are constructed to agree with the PDE, so they do not really model anything sensible. This system only serves to prove a point about non-existence, but is not in itself very interesting otherwise. An actually interesting (and highly important) example comes from Hans Lewy [2]. It consists of a simple first order PDE with smooth coefficients that does not admit any smooth solutions in any neighbourhood of the space.

This example is interesting in that even the smoothness of the coefficients do not guarantee the existence of a solution. However, when the coefficients are analytic functions, the Cauchy–Kovalevskaya theorem guarantees local solvability of the PDE.

Let us now proceed by utilizing the noncharacteristic condition to give us local invertibility of the characteristics, and with that a theorem for local existence of a solution for the whole PDE.

To build a solution u from the characteristic equations, at the minimum near the boundary Γ, we need to apply all of the developed theory. Select a point $x^0 \in \Gamma$, and without loss of generality assume that the surface near x^0 is flat, i.e. $\{x_n = 0\}$ and so $x^0 \in \mathbb{R}^{n-1}$. Suppose further that we have an admissible triple (p^0, z^0, x^0) and noncharacteristic boundary data with it. Now by Lemma 1 there is a function $q(\cdot)$, such that $q(x^0) = p^0$ with the triple $(q(y), g(y), y)$ being admissible for all $y \in \Gamma$ sufficiently close to x^0. With any such point y we solve the characteristic ODE (1.2.3) subject to initial conditions (1.3.4), and write:

$$\begin{cases} p(s) = p(y,s) = p(y_1, ..., y_{n-1}, s) \\ z(s) = z(y,s) = z(y_1, ..., y_{n-1}, s) \\ x(s) = x(y,s) = x(y_1, ..., y_{n-1}, s) \end{cases}$$

to display the solutions' natural dependence on s and y (so naturally not only the dependence of "time", but also where that characteristic started from).

Lemma 2. *(Local invertibility). Assume that the noncharacteristic condition $F_{p_n}(p^0, z^0, x^0) \neq 0$ holds. Then there exist an open interval $I \subset \mathbb{R}$ containing 0, a neighbourhood W of x^0 in $\Gamma \subset \mathbb{R}^{n-1}$, and a neighbourhood V of x^0 in $\mathbb{R}^n$, such that for each $x \in V$ there exist unique $s \in I$, $y \in W$ such that*

$$x = x(y, s),$$

where the mappings $x \to s, y$ are C^2.

Proof. Since $x(x^0, 0) = x^0$, the Inverse Function Theorem gives the result, provided $\det Dx(x^0, 0) \neq 0$. For any $y \in \Gamma$ we have $x(y, 0) = (y, 0)$, and so if $i = 1, ..., n-1$,

$$x^j_{y_i}(x^0, 0) = \begin{cases} \delta_{ij}, & \text{for} \quad j = 1, .., n-1 \\ 0, & \text{for} \quad j = n. \end{cases}$$

We also have from the characteristic equation (1.2.3)(c), that

$$x^j_s(x^0, 0) = F_{p_j}(p^0, z^0, x^0).$$

Now we can write the Jacobian of x at the point $(x^0, 0)$ (x only depends on y and s) to prove the unique solvability of it with regard to the other variables of F:

$$Dx(x^0,0) = \begin{pmatrix} 1 & \dots & 0 & F_{p_1}(p^0,z^0,x^0) \\ \vdots & \ddots & \vdots & \vdots \\ 0 & \dots & 1 & \vdots \\ 0 & \dots & 0 & F_{p_n}(p^0,z^0,x^0) \end{pmatrix}_{n\times n}.$$

Now again we have $\det Dx(x^0,0) = F_{p_n}(p^0,z^0,x^0) \neq 0$ as per the assumed noncharacteristic condition (1.3.6). An application of the Inverse Function Theorem now gives the result. □

The theorem asserts that the solvability of the function $u \in C^2$ will be given on V, and as such we have that the characteristics will not intersect in V due to their uniqueness. Notice also that the neighbourhood W need not be the whole boundary Γ, but can be instead chosen as the smallest subset of it so that we can get full solutions in V.

Now since $y \in \Gamma$ is close to x^0, we can have the solution depending on both s and y. Identifying the parameter s as time, it indeed makes sense to talk about how the unique characteristic curves emanate from different points of the boundary without crossing as time elapses. That is in the view of Lemma 2 we can locally uniquely solve the equation for each $x \in V$

$$\begin{cases} x = x(y,s) \\ \text{for} \quad y = y(x),\ s = s(x). \end{cases} \tag{1.4.3}$$

And so it is justified to set

$$\begin{cases} u(x) \coloneqq z(y(x),s(x)) \\ p(x) \coloneqq p(y(x),s(x)) \end{cases} \tag{1.4.4}$$

for $x \in V$ and s, y as in (1.4.3) (notice how they only explicitly depend on the specific curve x).

Now for the main result: we can locally weave the solutions of the characteristic ODE into a solution of the PDE.

Theorem 3. *(Local existence theorem). The function u defined above is C^2 and solves the PDE*

$$F(Du(x), u(x), x) = 0, \quad x \in V$$

with the boundary condition

$$u(x) = g(x), \quad x \in \Gamma \cap V.$$

Proof. Fix $y \in \Gamma$ close to x^0 and solve the characteristic ODE (1.2.3) subject to the initial conditions (1.3.4), for $p(s) = p(y,s)$, $z(s) = z(y,s)$ and $x(s) = x(y,s)$. For such a point y, we claim that

$$f(y,s) \coloneqq F(p(y,s), z(y,s), x(y,s)) = 0, \quad s \in I,$$

where $I \subset \mathbb{R}$ is as in Lemma 2. Firstly by the compatibility condition (1.3.5):

$$f(y,0) = F(p(y,0), z(y,0), x(y,0)) = F(q(y), g(y), y) = 0.$$

Secondly by the definition of f we have that:

$$f_s(y,s) = \sum_{j=1}^{n} F_{p_j} \dot{p}^j + F_z \dot{z} + \sum_{j=1}^{n} F_{x_j} \dot{x}^j,$$

and so by plugging in the identities (1.2.3), we get that

$$f_s(y,s) = 0.$$

The derivative of the function f with respect to the parameter s is 0, so it's a constant function of s. Furthermore since it attains a value 0 at all of the points $(y,0)$, we can conclude that $f(y,s) = 0$, for all $s \in I$.

By Lemma 2 and (1.4.3)-(1.4.4), we have

$$F(p(x), u(x), x) = 0, \quad x \in V.$$

So the last thing that needs to hold is

$$p(x) = Du(x), \quad x \in V.$$

For this let us prove some preliminary results from which the identity then immediately follows. We claim that for $s \in I$ and $y \in W$ we have:

$$z_s(y,s) = \sum_{j=1}^{n} p^j(y,s) x_s^j(y,s) \tag{1.4.5}$$

and

$$z_{y_i}(y,s) = \sum_{j=1}^{n} p^j(y,s) x_{y_i}^j(y,s), \quad i = 1, ..., n-1. \tag{1.4.6}$$

The first identity is immediate from (1.2.3)(b)-(c). For the second identity, fix $y \in \Gamma$, $i \in \{1, ..., n-1\}$, and set

$$r^i(s) := z_{y_i}(y,s) - \sum_{j=1}^{n} p^j(y,s)x^j_{y_i}(y,s). \tag{1.4.7}$$

If we can prove that this expression equals 0, we prove the identity (1.4.6).

First note that by the compatibility conditions (1.3.5) we have that $r^i(0) = g_{x_i}(y) - q^i(y) = 0$. Taking the derivative of r and the partial derivative of expression (1.4.5) with respect to y_i gives us

$$\dot{r^i}(s) = z_{y_i s} - \sum_{j=1}^{n} \big(p^j_s x^j_{y_i} + p^j x^j_{y_i s}\big)$$

and

$$z_{y_i s} = \sum_{j=1}^{n} \big(p^j_{y_i} x^j_s + p^j x^j_{s y_i}\big),$$

since $z(y(x), s(x)) = u(x) \in C^2$. Substituting $z_{y_i s}$ above, we get:

$$\dot{r^i}(s) = \sum_{j=1}^{n} \big(p^j_{y_i} x^j_s - p^j_s x^j_{y_i}\big) = \sum_{j=1}^{n} \big(p^j_{y_i} F_{p_j} - \big(- F_{x_j} - F_z p^j\big) x^j_{y_i}\big),$$

by (1.2.3)(a) and (c). Now finally differentiate f with respect to y_i to get

$$\sum_{j=1}^{n} F_{p_j} p^j_{y_i} + F_z z_{y_i} + \sum_{j=1}^{n} F_{x_j} x^j_{y_i} = 0,$$

and apply this identity above to get:

$$\dot{r^i}(s) = F_z\big(\sum_{j=1}^{n} p^j x^j_{y_i} - z_{y_i}\big) = -F_z r^i(s).$$

Hence $r^i(\cdot)$ solves this linear ODE, with the initial condition $r^i(0) = 0$. Solving this gives us that $r^i(s) = 0$ for $s \in I$ and $i = \{1, ..., n-1\}$; verifying the second identity (1.4.6).

Let us now utilize the two verified identities (1.4.5) and (1.4.6) to prove the final statement $p(x) = Du(x)$. So in fact for $j = 1, ..., n$,

$$\begin{aligned} u_{x_j} = z_{x_j}(y(x), s(x)) &= z_s s_{x_j} + \sum_{i=1}^{n-1} z_{y_i} y^i_{x_j} \\ &= \big(\sum_{k=1}^{n} p^k x^k_s\big) s_{x_j} + \sum_{i=1}^{n-1} \big(\sum_{k=1}^{n} p^k x^k_{y_i}\big) y^i_{x_j} \end{aligned}$$

$$= \sum_{k=1}^{n} p^k \big(x_s^k s_{x_j} + \sum_{i=1}^{n-1} x_{y_i}^k y_{x_j}^i \big)$$
$$= \sum_{k=1}^{n} p^k x_{x_j}^k = \sum_{k=1}^{n} p^k \delta_{jk} = p^j,$$

where the identities (1.4.5) and (1.4.6) were used on the second line, and the chain rule on the third. This verifies the last statement and thus concludes the proof.

□

The developed theory solves the PDE only locally. Problems might occur trying to extend the solutions globally. For example, as stated before, it could be that the characteristic curve intersects the initial curve more than once developing possible problems with the initial-conditions; this is due to the fact that the characteristic equation is well-posed for a single initial condition.

This completes our theory on the method of characteristics. Next we will apply it to different types of PDE:s.

1.5 Characteristics of linear, quasilinear and nonlinear PDE:s

1.5.1 Linear case

A linear and homogeneous PDE has the general form:

$$F(Du, u, x) = b(x) \cdot Du(x) + c(x)u(x) = 0, \quad x \in U. \tag{1.5.1}$$

Using the notation developed, we have $F(p, z, x) = b(x) \cdot p + c(x)z$ and $D_p F = b(x)$. Thus the system of characteristic equations become:

$$\begin{cases} \dot{z}(s) = b(x(s)) \cdot p(s) = -c(x(s))z(s) \\ \dot{x}(s) = b(x(s)). \end{cases} \tag{1.5.2}$$

The noncharacteristic condition at a point $x^0 \in \Gamma$ becomes

$$b(x^0) \cdot v(x^0) \neq 0$$

and thus does not involve z^0 or p^0 at all. By specifying the boundary conditions, we can uniquely solve the equation (1.3.5) for $q(y)$ when $y \in \Gamma$ is near

x^0. Theorem 3 can now be applied to construct a unique solution of (1.5.1) along with the boundary conditions in some neighbourhood V containing x^0. By uniqueness of solutions of the initial-value problem of the ODE (1.5.2), the projected characteristics $x(\cdot)$ originating from distinct points on Γ cannot cross each other. Thus a very efficient and straightforward method for providing the solutions is provided in the linear case. One should however not forget that Theorem 3 only ensures a local solution near the boundary; whether the solution can be smoothly continued to all of U has to be determined case by case.

1.5.2 Quasilinear case

A quasilinear and homogeneous PDE has the general form:

$$F(Du, u, x) = b(x, u(x)) \cdot Du(x) + c(x, u(x)) = 0, \quad x \in U. \tag{1.5.3}$$

This becomes $F(p, z, x) = b(x, z) \cdot p + c(x, z)$ with $D_pF = b(x, z)$. Hence the characteristic equations:

$$\begin{cases} \dot{z}(s) = b(x(s), z(s)) \cdot p(s) = -c(x(s), z(s)) \\ \dot{x}(s) = b(x(s), z(s)), \end{cases} \tag{1.5.4}$$

are similar to the linear case with no integration required on all of the equations of the full system (again no p dependence). The non-characteristic condition at a point $x^0 \in \Gamma$ becomes

$$b(x^0, z^0) \cdot v(x^0) \neq 0,$$

where $z^0 = g(x^0)$; and thus does not involve p^0 at all. Yet again, by specifying the boundary conditions, we can uniquely solve the equation (1.3.5) for $q(y)$ if $y \in \Gamma$ is near x^0. Theorem 3 can now be applied to construct a unique solution of (1.5.3) along with the boundary conditions in some neighbourhood V containing x^0. Unlike in the linear case however, the projected characteristics emanating from distinct points in Γ may intersect outside V (the range of the local solutions). This likely means that our local solution will not exist within all of U. Let us have one of such possible examples.

Example 3. *The scalar conservation law is a quasilinear first-order PDE:*

$$\begin{cases} G(Du, u_t, u, x, t) = u_t + divF(u) = 0 \quad in\ U = \mathbb{R}^n \times (0, \infty) \\ \qquad\qquad\qquad\qquad\qquad\qquad u = g \quad on\ \Gamma = \mathbb{R}^n \times \{t = 0\}. \end{cases} \tag{1.5.5}$$

Here $F : \mathbb{R} \to \mathbb{R}^n$, $F = (F^1, ..., F^n)$, $t = x_{n+1}$ and "div" is the divergence with respect to the spatial variables $x = (x_1, ..., x_n)$. We can further rewrite

the divergence as $divF(u) = F'(u) \cdot Du$, *where* $Du = D_x u = (u_{x_1}, ..., u_{x_n})$. *As the variable* t *plays an important role in our equations, we modify the notation so that* $q = (p, p_{n+1})$ *and* $y = (x, t)$. *Now we have as per our new notation*

$$G(q, z, y) = p_{n+1} + F'(z) \cdot p$$

and consequently

$$D_q G = (F'(z), 1),\ D_y G = 0,\ D_z G = F''(z) \cdot p.$$

Since $G_{p_{n+1}} = 1 \neq 0$, *our noncharacteristic condition is satisfied at each point* $y^0 = (x^0, 0) \in \Gamma$, *which allows us to use Lemma 1. Furthermore the equations (1.5.4) become*

$$\begin{cases} \dot{x}^i(s) & = F^{i\prime}(z(s)) \quad i = 1, .., n \\ \dot{x}^{n+1}(s) & = 1 \\ \dot{z}(s) & = 0. \end{cases}$$

By integrating the second term we get that $x^{n+1}(s) = s$ *which agrees with our notation above; we can identify the parameter* s *with time* t. *From the last term we get* $z(s) = z^0 = g(x^0)$ *and so this further implies that*

$$x(s) = F'(g(x^0))s + x^0.$$

Now the projected characteristic $y(s) = (x(s), s) = (F'(g(x^0))s + x^0, s)$ *for* $s \geq 0$ *is a straight line, along which* u *is a constant.*

Suppose now that we choose another initial point $\Gamma \ni z^0 \neq x^0$ with $g(x^0) \neq g(z^0)$. It is then clear that there is a possibility for the projected characteristics to intersect each other at some time $t > 0$. Theorem 2 now tells us that $u \equiv g(x^0)$ and $u \equiv g(z^0)$ on the projected characteristics through x^0 and z^0 respectively; a contradiction. From this we infer that the initial-value problem (1.5.5) does not in general have a smooth solution, existing for all times $t > 0$. Indeed Theorem 3 only guarantees the existence of local solutions for a short time. In the next section we will extend this to all times $t > 0$ by weakening some of the assumptions. Such solutions are known as "weak" or "generalized" solutions.

Remark. *Since for a given* $x \in \mathbb{R}^n$ *and* $t > 0$ *we have* $s = t$, *it is possible to write the solution* $u(x(t), t) = z(t)$ *implicitly as* $u = g(x - tF'(u))$. *This provides a solution granted* $1 + tDg(x - tF'(u)) \cdot F''(u) \neq 0$ *(differentiate both sides of the implicit equation with respect to* u*). Let* $n = 1$, *now if* $F'' > 0$ *and* $g' < 0$, *it is easy to see that the condition will definitely be false at some point in time* $t > 0$. *This would indeed be a failure of the characteristic method to form a solution for all times* t *as discussed above.*

1.5.3 Fully nonlinear case

Let's finally have a well known fully nonlinear example that we will look into further in the next chapter.

Example 4. *The general Hamilton-Jacobi PDE has the form*

$$G(Du, u_t, u, x, t) = u_t + H(Du, x) = 0, \tag{1.5.6}$$

where $Du = D_x u = (u_{x_1}, ..., u_{x_n})$ *is the "spatial" gradient. As before, write the equation in a more cohesive manner with* $q = (p, p_{n+1})$, $y = (x, t)$, *so that*

$$G(q, z, y) = p_{n+1} + H(p, x).$$

Now

$$D_q G = (D_p H(p, x), 1),\ D_y G = (D_x H(p, x), 0),\ D_z G = 0,$$

and since $G_{p_{n+1}} = 1 \neq 0$, *our noncharacteristic condition is satisfied at each point* $y^0 = (x^0, 0) \in \Gamma$, *which allows us to use Lemma 1 once again. The characteristic equations become:*

$$\begin{cases} \dot{p}^i(s) & = -H_{x_i}(p(s), x(s)) \quad for \quad i = 1, ..., n \\ \dot{p}^{n+1}(s) & = 0 \\ \dot{z}(s) & = D_p H(p(s), x(s)) \cdot p(s) + p^{n+1} \\ \dot{x}^i(s) & = H_{p_i}(p(s), x(s)) \quad for \quad i = 1, ..., n \\ \dot{x}^{n+1}(s) & = 1, \end{cases}$$

or more compactly:

$$\begin{cases} \dot{p}(s) = -D_x H(p(s), x(s)) \\ \dot{z}(s) = D_p H(p(s), x(s)) \cdot p(s) - H(p(s), x(s)) \\ \dot{x}(s) = D_p H(p(s), x(s)). \end{cases} \tag{1.5.7}$$

Again solving for $z(\cdot)$ *becomes trivial once* $x(\cdot)$ *and* $y(\cdot)$ *have been solved, which is why the first and third equations of (1.5.7) are somewhat special: the Hamilton's equations.*

As in the previous example, the initial-value problem for the Hamilton-Jacobi equation does not in general enjoy a smooth solution u for all times $t > 0$. However through the characteristics, we gain qualitative information of the PDE by examining the flow of values through the solution. The Hamilton-Jacobi equation is a hugely important concept (with vast generalizations), which is why the next chapter will be devoted to its investigation.

2 The Hamilton-Jacobi equations

In the previous section we focused on creating the characteristic equations for first-order partial differential equations. These equations allowed us to evolve initial data along an $(n-1)$-dimensional surface into an n-dimensional solution surface. As shown, this is solution method is highly restricted as being available only locally: the characteristic ODE is not allowed to have crossing characteristics (non-uniqueness) or finite time blow up (solutions' non-continuous dependence of the initial conditions). In this section our goal is to find an appropriate generalized solution (weak solution) for the Hamilton-Jacobi equations. This weak solution will exist for all times $t > 0$, even after the method of characteristics has failed. The solution will in its core hold some variational principles which we shall discuss next.

2.1 Variational methods

Recall the initial value problem for the Hamilton-Jacobi equation:

$$\begin{cases} u_t + H(Du) = 0 & \text{in } \mathbb{R}^n \times (0, \infty) \\ u = g & \text{on } \mathbb{R}^n \times \{t = 0\}, \end{cases} \tag{2.1.1}$$

where the *Hamiltonian* $H : \mathbb{R}^n \to \mathbb{R}$ and the initial function $g : \mathbb{R}^n \to \mathbb{R}$ are given, $u : \mathbb{R}^n \times [0, \infty)$ is the unknown, $u = u(x, t)$, and $Du = D_x u = (u_{x_1}, ..., u_{x_n})$. With this problem we associate the Hamilton's ODE

$$\begin{cases} \dot{p} = -D_x H(p, x) \\ \dot{x} = D_p H(p, x) \end{cases} \tag{2.1.2}$$

we acquired by the method of characteristics. Notice the how the Hamiltonian depends on both p and x, and how the ODE for $\dot{z}$ is not needed since it can be determined by these two equations alone.

Let us introduce a smooth function $L : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ called the *Lagrangian*, for which

$$L = L(v, x) = L(v_1, ..., v_n, x_1, ..., x_n) \quad v, x \in \mathbb{R}^n$$

and

$$\begin{cases} D_v L = (L_{v_1}, ..., L_{v_n}) \\ D_x L = (L_{x_1}, ..., L_{x_n}). \end{cases}$$

This is, in a "physical" sense, a function of generalized coordinates, their time derivatives, and time itself, that encodes the dynamics of the system.

An approach to the study of physical systems is the *principle of least action*. We can imagine an infinite dimensional space of "paths" a system might take, and in this space the "real" path we end up measuring. We will define such "real" path as minimizing a certain quantity, and satisfying certain equations (Euler-Lagrange equations to be exact).

Fix two points $x, y \in \mathbb{R}^n$ and a time $t > 0$. We now define the *action* functional

$$I[w(\cdot)] \coloneqq \int_0^t L(\dot{w}(s), w(s))\, ds, \tag{2.1.3}$$

for functions $w(\cdot) = (w^1(\cdot), ..., w^n(\cdot))$ belonging to the *admissible class*

$$A = \{w(\cdot) \in C^2([0,t]; \mathbb{R}^n) \,|\, w(0) = y, w(t) = x\}.$$

Expression (2.1.3) is well defined since L is assumed to be smooth and $w \in A$. In this manner both the action and Lagrangian contain the dynamics of the system for all times from 0 to t. Notice how in the definition of the action functional the name of the variable v has been substituted with $\dot{w}(s)$, and x with $w(s)$.

The premise for admissible class of curves is very simple: a curve w is in the class if its twice continuously differentiable and if it starts at a point y at time 0, and reaches the point x at time t. We seek a curve $x \in A$ that minimizes I over A. That is, the basic problem in the *calculus of variations* is to find a curve $x(\cdot) \in A$ satisfying

$$I[x(\cdot)] = \min_{w(\cdot) \in A} I[w(\cdot)]. \tag{2.1.4}$$

Example 5. *An elementary example of such a problem is finding a curve of minimum length between two points (a, b) and (c, d) in the plane $\mathbb{R}^2$. Assume that the minimizing curve is given as the graph of a function $w = f(s)$. We know that the length of such a curve is*

$$I[w] = \int_a^c \sqrt{1 + w_s^2}\, ds = \int_a^c \sqrt{1 + f'(s)^2}\, ds.$$

The problem is then to minimize this functional over the space of differentiable functions $w = f(s)$ such that, for example, $f(a) = b$ and $f(c) = d$.

Remark. *A more general example of the principle of least action would be to find a geodesic between two points on a manifold. This would happen by finding the minimum of the "distance functional" which seeks to minimize the distance on a set of all smooth paths between the two points on the manifold.*

Next we will work out some properties of a curve $x(\cdot)$ satisfying (2.1.4) (its existence shall be assumed).

Theorem 4. *(Euler-Lagrange equations). The function $x(\cdot)$ solves the system of Euler-Lagrange equations*

$$-\frac{d}{ds}(D_v L(\dot{x}(s), x(s))) + D_x L(\dot{x}(s), x(s)) = 0, \quad \textit{where } 0 \le s \le t \tag{2.1.5}$$

This is a vector equation, consisting of n coupled second-order equations. The following proof will show that any minimizer $x(\cdot) \in A$ of $I[\cdot]$ solves the Euler-Lagrange system of ODE. It can of course be, that some $x(\cdot) \in A$ that is not a minimizer, solves the Euler-Lagrange equations. Such curves are called *critical points* of $I[\cdot]$. Obviously then every minimizer is a critical point, but a critical point need not be a minimizer.

Proof. Let $y : [0,t] \to \mathbb{R}^n$, $y(\cdot) = (y^1(\cdot), ..., y^n(\cdot))$ be a smooth function with $y(0) = y(t) = 0$. Define a path

$$w(\cdot) \coloneqq x(\cdot) + cy(\cdot), \quad c \in \mathbb{R}. \tag{2.1.6}$$

Clearly $w(\cdot) \in A$ and $I[x(\cdot)] \le I[w(\cdot)]$, and thus the real-valued function $i(c) \coloneqq I[w(\cdot)] = I[x(\cdot) + cy(\cdot)]$ has a minimum at $c = 0$; consequently

$$\frac{d}{dc} i(0) = i'(0) = 0,$$

provided $i'(0)$ exists.

Now by definition of the action functional

$$i(c) = \int_0^t L(\dot{x}(s) + c\dot{y}(s), x(s) + cy(s))\, ds$$

and so by the smoothness of L we have

$$i'(c) = \int_0^t \sum_{i=1}^n L_{v_i}(\dot{x} + c\dot{y}, x + cy)\dot{y}^i + L_{x_i}(\dot{x} + c\dot{y}, x + cy)y^i\, ds.$$

Set $c = 0$ so that

$$0 = i'(0) = \int_0^t \sum_{i=1}^n L_{v_i}(\dot{x}, x)\dot{y}^i + L_{x_i}(\dot{x}, x)y^i\, ds.$$

Integrating by parts we have that

$$\int_0^t L_{v_i}(\dot{x}, x)\dot{y}^i\, ds = \left[L_{v_i}(\dot{x}, x)y^i\right]_0^t - \int_0^t \frac{d}{ds} L_{v_i}(\dot{x}, x)\dot{y}^i\, ds,$$

and since $y(0) = y(t) = 0$, we get

$$\int_0^t L_{v_i}(\dot{x}, x)\dot{y}^i \, ds = -\int_0^t \frac{d}{ds} L_{v_i}(\dot{x}, x)\dot{y}^i \, ds.$$

Substituting this above, we acquire the identity

$$0 = \sum_{i=1}^{n} \int_0^t \Big[-\frac{d}{ds}\big(L_{v_i}(\dot{x}, x)\big) + L_{x_i}(\dot{x}, x)\Big] y^i \, ds.$$

As this identity is valid for all smooth functions y satisfying the boundary conditions $y(0) = y(t) = 0$, it must hold that

$$-\frac{d}{ds}\big(L_{v_i}(\dot{x}, x)\big) + L_{x_i}(\dot{x}, x) = 0, \quad \text{for} \quad 0 \le s \le t, \ i = 1, .., n.$$

□

Example 6. *Euler-Lagrange equations of Example 5. In this case we have* $L(f'(s), f(s), s) = \sqrt{1 + f'(s)^2}$, *with*

$$D_f L = 0 \quad \textit{and} \quad D_{f'} L = \frac{f'}{\sqrt{1 + f'(s)^2}}.$$

Using the Euler-Lagrange equations we obtain

$$\frac{d}{ds}\frac{f'}{\sqrt{1 + f'(s)^2}} = 0 \implies \frac{f'}{\sqrt{1 + f'(s)^2}} = C$$

for some constant $0 < C^2 < 1$. *Solving this we get:*

$$f'(s) = \frac{C}{\sqrt{1 - C^2}} \coloneqq A \implies f(s) = As + B,$$

for constants $A = \frac{d-b}{c-a}$ *and* $B = \frac{cb-ad}{c-a}$. *The graph is thus a straight line, i.e. the shortest distance between two points in the plane* $\mathbb{R}^2$ *is a straight line.*

Another notable example of a Euler-Lagrange equation is Newton's second law of motion.

Example 7. *Consider the Lagrangian* $L(v, x) = \frac{1}{2}m|v|^2 - \phi(x)$, *(the difference between the kinetic and potential energies) where* $m > 0$. *Then by the Euler-Lagrange equations we have:*

$$\frac{d}{ds} D_v L(\dot{x}(s), x(s)) - D_x L(\dot{x}(s), x(s)) = m\ddot{x}(s) - f(x(s)) = 0.$$

This describes a particle with mass m *moving in a force field* $f \coloneqq -D\phi$ *generated by the potential* ϕ.

2.2 Hamilton's ODE

Next we introduce the Hamiltonian and intimately connect it with the Lagrangian. This will give us a tool to convert the Euler-Lagrange equations into Hamilton's equations; a system of n second-order ODE to a system of $2n$ system of first-order ODE. This drop in the higher order derivatives may not give an advantage on what comes to solving it, but important theoretical results can be derived because of the nearly symmetric roles of the independent variables: coordinates and momenta.

Assume that the C^2 function $x(\cdot)$ is a critical point of the action functional, and thus solves the Euler-Lagrange equations. We define a *generalized momentum* $p(\cdot)$ corresponding to the position $x(\cdot)$ and velocity $\dot{x}(\cdot)$ by:

$$p(s) \coloneqq D_v L(\dot{x}(s), x(s)), \quad 0 \leq s \leq t. \tag{2.2.1}$$

Regarding this we need to still make one important assumption:

$$\begin{cases} \text{Suppose for all } x, p \in \mathbb{R}^n \text{ that the equation} \\ p = D_v L(v, x) \text{ can be unquely solved for } v \\ \text{as a smooth function of } p \text{ and } x : v = v(p, x). \end{cases} \tag{2.2.2}$$

Definition 6. *The Hamiltonian H associated with the Lagrangian L is*

$$H(p, x) \coloneqq p \cdot v(p, x) - L(v(p, x), x), \quad p, x \in \mathbb{R}^n, \tag{2.2.3}$$

where the function $v(\cdot)$ is defined implicitly by (2.2.2).

Example 8. *Returning to our previous example, Newton's second law. If the equation $p = D_v L(v, x)$ can be uniquely solved for v as a smooth function of p and x, then we have $p = D_v L(v, x) = mv$, so that $v(p, x) = v = \frac{p}{m}$. The Hamiltonian corresponding to its Lagrangian is*

$$H(p, x) = D_v L(v, x) \cdot v - L(v(p, x), x) = \frac{1}{2m}|p|^2 + \phi(x).$$

The Hamiltonian is thus the sum of the kinetic and potential energies: the total energy of the system.

Theorem 5. *(Derivation of Hamilton's ODE). The functions $x(\cdot)$ and $p(\cdot)$ satisfy the coupled system of $2n$ first-order Hamilton's equations:*

$$\begin{cases} \dot{p}(s) = -D_x H(p(s), x(s)) \\ \dot{x}(s) = D_p H(p(s), x(s)) \end{cases} \tag{2.2.4}$$

for $0 \leq s \leq t$. Furthermore, the mapping $s \to H(p(s), x(s))$ is constant.

Remark. *The constancy of H with respect to the parameter s means that the trajectories lie on the contour lines $H(p,x) = C$. This can be viewed as the conservation of energy of the system it describes (time translation invariance). Thus the equations are particularly useful in identifying conserved quantities for mechanical systems; this being true even when the problem itself cannot be solved completely.*

Proof. From $p(s) = D_v L(\dot{x}(s), x(s))$, $p = D_v L(v,x)$ and $v = v(p,x)$, we get that $\dot{x}(s) = v(p(s), x(s))$. Denote $v(\cdot) = (v^1(\cdot), ..., v^n(\cdot))$ and compute the partial derivatives of H for $i = 1, .., n$:

$$H_{x_i}(p,x) = \sum_{k=1}^{n} p_k v^k_{x_i}(p,x) - L_{v_k}(v,x) v^k_{x_i}(p,x) - L_{x_i}(v,x) = -L_{x_i}(v,x)$$

and

$$H_{p_i}(p,x) = v^i(p,x) + \sum_{k=1}^{n} p_k v^k_{p_i}(p,x) - L_{v_k}(v,x) v^k_{p_i}(p,x) = v^i(p,x),$$

both by (2.2.2). Thus

$$H_{p_i}(p(s), x(s)) = v^i(p(s), x(s)) = \dot{x}^i(s)$$

and by the Euler-Lagrange equations

$$\begin{aligned} H_{x_i}(p(s), x(s)) &= -L_{x_i}(v(p(s), x(s)), x(s)) \\ &= -L_{x_i}(\dot{x}(s), x(s)) = -\frac{d}{ds}\Big(L_{v_i}(\dot{x}(s), x(s)) \Big) \\ &= -\dot{p}^i(s). \end{aligned}$$

For the last assertion simply take the derivative with respect to s and use the identities acquired above;

$$\frac{d}{ds} H(p(s), x(s)) = \sum_{i=1}^{n} H_{p_i} \dot{p}^i + H_{x_i} \dot{x}^i = \sum_{i=1}^{n} H_{p_i} \Big(- H_{x_i} \Big) + H_{x_i} \Big(H_{p_i} \Big) = 0.$$

□

What does the Hamilton's ODE mean in Example 8? Firstly, the time derivative of the generalized momentum p equals the Newtonian force; that is the first Hamilton equation means that the force equals the negative gradient of the potential energy. The time derivative of x is the velocity, and so the second Hamilton equation means that the particle's velocity equals the

derivative of its kinetic energy with respect to its momentum. Even though Hamiltonian mechanics can be used to describe simple systems such as a bouncing ball or an oscillating spring in which the energy changes from kinetic to potential and back again over time, its true potential is shown in more complex dynamical systems. A fair example would be something with many degrees of freedom (which complicates the systems' time evolution), such as planetary orbits in celestial mechanics. Very generally speaking, you don't have the system fully described if you haven't got the Hamiltonian.

2.3 Duality and the Hopf-Lax formula

Now we connect the Hamilton-Jacobi PDE with the minimization problem from calculus of variations. Let us return to the definition of the Hamiltonian and make the simplification: $H(p,x) = H(p) = H$. We will introduce a transformation that shows the duality of H and L.

2.3.1 Legendre transform

The Legendre transform is an involutive (its own inverse) transform on the real-valued convex functions of one real variable. For sufficiently smooth functions it tells that a functions' and its Legendre transforms' first derivatives are each others inverse functions. Let's suppose that the Lagrangian $L : \mathbb{R}^n \to \mathbb{R}^n$ satisfies: the mapping

$$v \to L(v)$$

is convex and

$$\lim_{|v|\to\infty} \frac{L(v)}{|v|} = \infty.$$

The continuity of L automatically follows from its convexity. A function L is called *superlinear* if the second condition applies.

Definition 7. *The Legendre transform of L is*

$$L^*(p) = \sup_{v\in\mathbb{R}^n} \{p \cdot v - L(v)\}, \quad p \in \mathbb{R}^n. \tag{2.3.1}$$

Remark. *The Legendre transform is a special case of a functions' convex conjugate, also known as Legendre-Fenchel transform.*

First note that the supremum is actually a maximum.

Proof. Choosing $v = 0$ gives us the inequality $L^*(p) \geq -L(0)$ for all $p \in \mathbb{R}^n$. Furthermore the superlinearity of L implies that for each $p \in \mathbb{R}^n$ there is an $M > 0$ such that for all $v \in \mathbb{R}^n$ for which $|v| > M$, we have

$$p \cdot v - L(v) < -L(0).$$

Thus we can ignore such v and conclude that

$$L^*(p) = \sup_{|v| \leq M} \{p \cdot v - L(v)\}, \;\; p \in \mathbb{R}^n.$$

The mapping $v \to p \cdot v - L(v)$ is continuous and the set $|v| \leq M$ compact so by the extreme value theorem the function achieves a maximum at some point $v = v^*$.

□

Denoting v^* as the maximum of the Legendre transform and assuming differentiability of L at it, we have that $p = DL(v^*)$. This means that the equation $p = DL(v)$ can be solved for v in terms of p, $v^* = v(p)$. Whether this the solution is unique or not, we have that

$$L^*(p) = p \cdot v(p) - L(v(p)).$$

However this is exactly the definition (2.2.3) for the Hamiltonian without the x dependence.

The Hamiltonian is thus the Legendre transform of the Lagrangian when holding v fixed and defining the p as the dual variable. We henceforth write

$$H = L^*. \tag{2.3.2}$$

Now we know how to obtain the Hamiltonian H from the Lagrangian L but what about the converse: given H, how do we compute L?

Theorem 6. *Assume that L is convex and superlinear, and define H by (2.3.1) and (2.3.2). Then the mapping*

$$p \to H(p)$$

is convex and

$$\lim_{|p| \to \infty} \frac{H(p)}{|p|} = \infty,$$

with

$$L = H^*.$$

Proof omitted. Thus the Lagrangian is the Legendre transform of the Hamiltonian. We have our main result:

$$L = H^*,\ H = L^*.$$

We call H and L dual convex functions. Assuming that only H or L is convex is enough for the duality result since it can be shown that applying the Legendre transform to a convex function again gives a convex function.

Remark. *Both approaches, the Hamiltonian and Lagrangian, give the same equations for the same generalized momentum. Which one to use is based mostly on circumstance: Lagrangian is based on Riemannian geometry and Hamiltonian on sympletic geometry.*

2.3.2 Hopf-Lax formula

Let us update the characteristic equations of the Hamilton-Jacobi equations (1.5.7). Since $H(p, x) = H(p)$, we have that

$$\begin{cases} \dot{p}(s) = 0 \\ \dot{z}(s) = D_p H(p) \cdot p(s) - H(p) \\ \dot{x}(s) = D_p H(p). \end{cases} \tag{2.3.3}$$

Inserting $\dot{x}$ to the expression for $\dot{z}$ and using the Legendre transform further gives us

$$\dot{z} = \dot{x} \cdot p(s) - H(p) = L(\dot{x}).$$

The characteristics provide a smooth solution u for at least short times $t > 0$ so that $z(t) = u(x(t), t)$ and therefore by integrating $\dot{z}$ we get:

$$u(x, t) = \int_0^t L(\dot{x})\, ds + g(x(0)).$$

We have thus turned the characteristic equations of the Hamilton-Jacobi PDE to a variational problem of the Lagrangian. This however should not come as a surprise since the calculus of variations problem led to Hamilton's ODE for the associated Hamiltonian (those being part of the characteristic equations of the Hamilton-Jacobi PDE).

Let us try to to somehow extend the solutions to times further than where the smoothness ends. Given $x \in \mathbb{R}^n$ and $t > 0$, our goal is to minimize the modified action

$$\int_0^t L(\dot{w}(s))\, ds + g(w(0))$$

among curves $w(\cdot)$ satisfying $w(t) = x$. We define

$$u(x,t) \coloneqq \inf\left\{ \int_0^t L(\dot{w}(s))\, ds + g(w(0)) \mid w(t) = x \right\}, \tag{2.3.4}$$

where the infimum is taken over all C^1 functions $w(\cdot)$. So in what sense does u defined by (2.3.4) actually solve the Hamilton-Jacobi PDE (2.1.1). Recall that we are assuming that H is smooth, convex and superlinear. Furthermore we assume that the initial condition satisfies:

$$g : \mathbb{R}^n \to \mathbb{R}^n \quad \text{is Lipschitz continuous,}$$

with

$$\mathrm{Lip}(g) \coloneqq \sup_{x,y\in\mathbb{R}^n, x\neq y} \left\{ \frac{|g(x) - g(y)|}{|x-y|} \right\} < \infty.$$

We start by simplifying the expression (2.3.4). The vector p is constant along the characteristic curve, and so $\dot{x} = D_p H$ is a constant vector, and therefore the projected characteristics $x(t)$ are straight lines. Thus if $x(0) = y$ and $x(t) = x$, we must have $\dot{x} = \frac{x-y}{t}$ and as a consequence:

$$\dot{z} = L(\dot{x}) = L\left(\frac{x-y}{t}\right) \implies z(t) = z(0) + tL\left(\frac{x-y}{t}\right) = g(y) + tL\left(\frac{x-y}{t}\right).$$

The only unknown term in this expression is y, so the problem turns into minimizing $z(t)$. The intuition behind this is the following: imagine that instead of starting from time $t = 0$, we start at say $t = -1$. All of our trajectories now start from $t = -1$, pass y at $t = 0$, and end at x at some defined $t > 0$. The initial function g can be thought of tracking the work done from point $t = -1$ to point $t = 0$; so now the correct trajectory is the one minimizing this work.

Theorem 7. *(Hopf-Lax formula). If $x \in \mathbb{R}^n$ and $t > 0$, then the solution $u = u(x,t)$ of the minimization problem (2.3.4) is*

$$u(x,t) = \min_{y\in\mathbb{R}^n} \left\{ tL\left(\frac{x-y}{t}\right) + g(y) \right\}. \tag{2.3.5}$$

Proof. Fix any $y \in \mathbb{R}^n$ and define $w(s) \coloneqq y + \frac{s}{t}(x-y)$, where $0 \le s \le t$. Then (2.3.4) implies

$$u(x,t) \le \int_0^t L(\dot{w})\, ds + g(y) = tL\left(\frac{x-y}{t}\right) + g(y),$$

which implies that

$$u(x,t) \leq \inf_{y\in\mathbb{R}^n}\left\{tL\left(\frac{x-y}{t}\right)+g(y)\right\}.$$

For the other direction let $w(\cdot)$ be any C^1 function satisfying $w(t)=x$. Since L is convex, by Jensen's inequality we have that

$$L\left(\frac{1}{t}\int_0^t \dot{w}(s)\,ds\right) \leq \frac{1}{t}\int_0^t L(\dot{w}(s))\,ds.$$

Defining $y=w(0)$ and adding $g(y)$ to both sides, we get

$$tL\left(\frac{x-y}{t}\right)+g(y) \leq \int_0^t L(\dot{w}(s))\,ds + g(y).$$

Taking infimum on both sides now gives us

$$\inf_{y\in\mathbb{R}^n}\left\{tL\left(\frac{x-y}{t}\right)+g(y)\right\} \leq u(x,t).$$

Combining the two estimates now gives us (2.3.5) for infimum.

The final claim is then that in the expression for u, the infimum can be replaced by an actual minimum. That is:

$$u(x,t)=\inf_{y\in\mathbb{R}}\left\{tL\left(\frac{x-y}{t}\right)+g(y)\right\}=\min_{y\in\mathbb{R}}\left\{tL\left(\frac{x-y}{t}\right)+g(y)\right\}.$$

Fix t and x, and define function h by

$$h(y)=tL\left(\frac{x-y}{t}\right)+g(y), \quad y\in\mathbb{R}^n.$$

The function h is continuous as a sum of continuous functions, and since g is Lipschitz, we have the estimate:

$$\begin{aligned} h(y) &\geq tL\left(\frac{x-y}{t}\right) - ||g||_{Lip}|x-y| - |g(x)| \\ &= |x-y|\left(\frac{L(\frac{x-y}{t})}{\frac{|x-y|}{t}} - ||g||_{Lip} - \frac{|g(x)|}{|x-y|}\right). \end{aligned}$$

Let $|y|\to\infty$ and notice that $\frac{L(\frac{x-y}{t})}{\frac{|x-y|}{t}}=\infty$ by superlinearity, with $\frac{|g(x)|}{|x-y|}=0$ clearly. This implies that $h(y)\to\infty$, and that there exists an $M>0$ such

that if $|y| > M$, then $h(y) > u(x,t) + \lambda$, where $\lambda > 0$ is fixed. Now by the extreme value theorem h attains its minimum on $\bar{B}(0,M)$ at some point y_0, and so $\inf_{|y|\leq M} h(y) \geq h(y_0)$. By the definition of the infimum there exists a $y' \in \bar{B}(0,M)$ such that $h(y') < u(x,t) + \lambda/2$. Now:

$$\inf_{|y|>M} h(y) \geq u(x,t) + \lambda > u(x,t) + \lambda/2 > h(y') \geq h(y_0).$$

Combining both estimates we have that $u(x,t) \geq h(y_0)$, and hence the minimum is attained for every fixed x and t.

□

Remark. *In fact the Hopf-Lax formula (2.3.5) reads*

$$u(x,t) = \min_{y\in B(x,Rt)} \left\{ tL\left(\frac{x-y}{t}\right) + g(y) \right\},$$

for $R = \sup_{\mathbb{R}^n} |DH(Dg)|$ *and* $L = H^*$. *The minimizer* y *is thus always bounded with respect to the time* t, *and therefore shows the finite propagation speed for the Hamilton-Jacobi equations with convex Hamiltonian and Lipschitz continuous initial function* g. *This can be proven via the subdifferential of* H.

The explicit formula of the Hopf-Lax was possible because of the special structure in the Hamilton-Jacobi characteristic equations, namely the Hamilton's equations, that allowed us to turn the ODE into a variational problem. Because of this explicitness, the formula has some useful properties. Our ultimate goal is to show that the formula produces a reasonably defined weak solution of the initial-value problem for the Hamilton-Jacobi equation (2.1.1), since it does not in general have a smooth solution u lasting for all times $t > 0$.

Lemma 3. *(Functional identity). For each* $x \in \mathbb{R}^n$ *and* $0 \leq s \leq t$, *we have*

$$u(x,t) = \min_{y\in R^n} \left\{ (t-s)L\left(\frac{x-y}{t-s}\right) + u(y,s) \right\}. \tag{2.3.6}$$

Proof omitted. This identity tells us that to compute $u(\cdot,t))$, we can just calculate u at time s and then use $u(\cdot,s))$ as the initial condition on the remaining time interval $[s,t]$.

Lemma 4. *The function* u *is Lipschitz continuous in* $\mathbb{R}^n \times [0,\infty)$, *and*

$$u = g \quad \text{on } \mathbb{R}^n \times \{t = 0\}.$$

Proof omitted. Now Rademacher's theorem says that a Lipschitz function is differentiable almost everywhere, so u defined by the Hopf-Lax formula (2.3.5) is differentiable a.e. $(x,t) \in \mathbb{R}^n \times (0,\infty)$. The proof of Rademacher's theorem can be found as a part of Sobolev-space theory from pages 280-281 in [1]. Now we are ready to show that the Hopf-Lax formula actually provides us a solution to the Hamilton-Jacobi PDE wherever u is differentiable.

Theorem 8. *Suppose $x \in \mathbb{R}^n$, $t > 0$, and u defined by the Hopf-Lax formula (2.3.5) is differentiable at a point $(x,t) \in \mathbb{R}^n \times (0,\infty)$. Then*

$$u_t(x,t) + H(Du(x,t)) = 0.$$

Proof. Fix $v \in \mathbb{R}^n$, $h > 0$. By Lemma 3 we have

$$u(x+hv, t+h) = \min_{y \in \mathbb{R}^n} \left\{ hL\left(\frac{x+hv-y}{h}\right) + u(y,t) \right\} \leq hL(v) + u(x,t),$$

where we chose $x = y$ (not necessarily the minimizer). This gives us the difference quotient

$$\frac{u(x+hv, t+h) - u(x,t)}{h} \leq L(v).$$

Let $h \to 0^+$ so that the directional derivative becomes

$$v \cdot Du(x,t) + u_t(x,t) \leq L(v).$$

This inequality is valid for all $v \in \mathbb{R}^n$, so by $H = L^*$ we have that:

$$u_t(x,t) + H(Du(x,t)) = u_t(x,t) + \max_{v \in \mathbb{R}^n} \{ v \cdot Du(x,t) - L(v) \} \leq 0.$$

For the other direction we choose a $z \in \mathbb{R}^n$ such that it is the minimizer of the Hopf-Lax formula: $u(x,t) = tL(\frac{x-z}{t}) + g(z)$. Fix $h > 0$ and set $s = t - h$ and $y = \frac{s}{t}x + (1 - \frac{s}{t})z$, so that $\frac{x-z}{t} = \frac{y-z}{s}$, and thus:

$$\begin{aligned} u(x,t) - u(y,s) &\geq tL\left(\frac{x-z}{t}\right) + g(z) - \left[sL\left(\frac{y-z}{s}\right) + g(z) \right] \\ &= (t-s)L\left(\frac{x-z}{t}\right). \end{aligned}$$

This again gives us a difference quotient by reorganization:

$$\frac{u(x,t) - u((1-\frac{h}{t})x + \frac{h}{t}z, t-h)}{h} \geq L\left(\frac{x-z}{t}\right).$$

Since $(1-\frac{h}{t})x+\frac{h}{t}z = x-h(\frac{x-z}{t})$, we can let $h \to 0^+$ to achieve

$$\frac{x-z}{t}\cdot Du(x,t)+u_t(x,t) \geq L\left(\frac{x-z}{t}\right),$$

and so

$$\begin{aligned} u_t(x,t)+H(Du(x,t)) &= u_t(x,t)+\max_{v\in R^n}\{v\cdot Du(x,t)-L(v)\} \\ &\geq u_t(x,t)+\frac{x-z}{t}\cdot Du(x,t)-L\left(\frac{x-z}{t}\right) \\ &\geq 0. \end{aligned}$$

This inequality completes the proof. □

From this follows a useful lemma for comparing solutions based on their initial functions. The solution u depends monotonically on g: if the initial-data g is increased pointwise, then so is the solution u.

Lemma 5. *(L^∞-contraction inequality). Let u^1 and u^2 be two solutions of the initial value problems*

$$\begin{cases} u_t^i + H(Du^i) = 0 & \text{a.e. in } \mathbb{R}^n\times(0,\infty) \\ u^i = g^i & \text{on } \mathbb{R}^n\times\{t=0\} \ (i=1,2), \end{cases}$$

given by the Hopf-Lax formula. Then we have the following inequality

$$\sup_{\mathbb{R}^n}|u^1(\cdot,t)-u^2(\cdot,t)| \leq \sup_{\mathbb{R}^n}|g^1-g^2|, \quad \text{for } t>0.$$

Furthermore if $g^2 \leq g^1$, then $u^2 \leq u^1$.

Proof. By the Hopf-Lax formula we have:

$$u^1(x,t) \leq \left\{tL\left(\frac{x-y}{t}\right)+g^1(y)\right\} \text{ and } u^2(x,t) \leq \left\{tL\left(\frac{x-y}{t}\right)+g^2(y)\right\},$$

where equalities hold with y_1 and y_2 respectively. Especially we have

$$u^1(x,t) \leq \left\{tL\left(\frac{x-y_2}{t}\right)+g^1(y_2)\right\} \text{ and } u^2(x,t) \leq \left\{tL\left(\frac{x-y_1}{t}\right)+g^2(y_1)\right\}.$$

From these we have the estimates $u^1(x,t)-u^2(x,t) \leq g^1(y_2)-g^2(y_2)$ and $u^1(x,t)-u^2(x,t) \geq g^1(y_1)-g^2(y_1)$. Taking supremum over both inequalities gives us the first claim.

The second claim is immediate from our second inequality above combined with the assumption $g^2 \leq g^1$. □

Let us summarize the section with the following theorem:

Theorem 9. *(Hopf-Lax formula as a solution). The function u defined by the Hopf-Lax formula (2.3.5) is Lipschitz continuous, differentiable a.e. in $\mathbb{R}^n \times (0,\infty)$, and solves the initial-value problem*

$$\begin{cases} u_t + H(Du) = 0 & \text{a.e. in } \mathbb{R}^n \times (0,\infty) \\ u = g & \text{on } \mathbb{R}^n \times \{t = 0\}. \end{cases} \tag{2.3.7}$$

Remark. *By using the Legendre transform on L, we can write the Hopf-Lax formula (2.3.5) as*

$$u(x,t) = \min_{y \in \mathbb{R}^n} \max_{z \in \mathbb{R}^n} \{z \cdot (x-y) - tH(z) + g(y)\}.$$

For each fixed y, z this function solves the PDE (2.1.1). Thus the Hopf-Lax formula builds a solution of (2.1.1) by taking appropriate two parameter envelopes of these functions using minima and maxima. This is evident since for a fixed pair y, z this expression is exactly the complete integral of Example 1 (with the choices $a = z$ and $b = g(y) - z \cdot y$).

2.4 Weak solutions and their uniqueness

Looking at the theorem above, it is tempting to define a suitable weak solution u in a similar manner. This definition would however produce problems with uniqueness of the solutions as the next example shows.

Example 9. *Consider the initial value problem*

$$\begin{cases} u_t + |u_x|^2 = 0 & \text{in } \mathbb{R} \times (0,\infty) \\ u = 0 & \text{on } \mathbb{R} \times \{t = 0\}. \end{cases} \tag{2.4.1}$$

Besides the trivial solution $u_0(x,t) \equiv 0$; we have

$$u_1(x,t) = \begin{cases} 0 & \text{if } |x| \geq t \\ x - t & \text{if } 0 \leq x \leq t \\ -x - t & \text{if } -t \leq x \leq 0 \end{cases}$$

which is Lipschitz continuous and solves the PDE everywhere except on the lines $x = 0, \pm t$. There are actually infinitely many Lipschitz functions satisfying (2.4.1) as can be seen from this family of solutions:

$$u_a(x,t) = \begin{cases} 0 & \text{if } |x| \geq t \\ ax - a^2 t & \text{if } 0 \leq x \leq t \\ -ax - a^2 t & \text{if } -t \leq x \leq 0, \end{cases}$$

where $a \in \mathbb{R}$.

Hence for uniqueness we must require more than just solvability of the PDE a.e and Lipschitz continuity of g. The next lemma shows that u will inherit a form of one-sided second-derivative estimate from the initial function g, granted that g be semiconcave. Semiconcavity will turn out to be a sufficient condition for the uniqueness to hold.

Lemma 6. *(Semiconcavity). Suppose there exist a constant C such that*

$$g(x+y) - 2g(x) + g(x-z) \leq C|z|^2 \tag{2.4.2}$$

for all $x, z \in \mathbb{R}^n$. Let u be defined by the Hopf-Lax formula (2.3.5). Then

$$u(x+z,t) - 2u(x,t) + u(x-z,t) \leq C|z|^2$$

for all $x, z \in \mathbb{R}^n$, $t > 0$.

Proof omitted. A function g is called semiconcave provided that (2.4.2) holds. If we were to assume that g was twice continuously differentiable with $\sup_{\mathbb{R}^n} |D^2 g| < \infty$, then the semiconcavity condition would automatically hold. Another characterisation would be that g is semiconcave if and only if the mapping $x \to g(x) - \frac{C}{2}|x|^2$ is concave for some constant $C \in \mathbb{R}$. Proof for this can be found in [3] as Proposition 1.1.3.

Requiring uniform convexity from the Hamiltonian H instead of semiconcavity from g leads us to the same conclusion.

Definition 8. *A convex C^2 function $H : \mathbb{R}^n \to \mathbb{R}$ is called uniformly convex with a constant $\theta > 0$ if*

$$\sum_{i,j=1}^{n} H_{p_i p_j}(p)\xi_i \xi_j \geq \theta|\xi|^2 \quad \text{for all } p, \xi \in \mathbb{R}^n. \tag{2.4.3}$$

Without g being semiconcave, the uniform convexity of H will force u to become semiconcave for $t > 0$ which is exactly what happened before with g. This regularizes the Hopf-Lax and provides uniqueness of the solution.

Lemma 7. *(Semiconcavity). Suppose that H is uniformly convex with a constant θ and u is defined by the Hopf-Lax formula (2.3.5). Then*

$$u(x+z,t) - 2u(x,t) + u(x-z,t) \leq \frac{1}{\theta t}|z|^2$$

for all $x, z \in \mathbb{R}^n$, $t > 0$.

Proof omitted. These semiconcavity conditions will ensure unique solutions from the Hopf-Lax formula.

Definition 9. *We say that a Lipschitz continuous function $u : \mathbb{R}^n \times [0,\infty) \to \mathbb{R}$ is a weak solution of the initial-value problem:*

$$\begin{cases} u_t + H(Du) = 0 & \text{in } \mathbb{R}^n \times (0,\infty) \\ u = g & \text{on } \mathbb{R}^n \times \{t = 0\} \end{cases} \tag{2.4.4}$$

provided

1. $u(x,0) = g(x)$ *for* $x \in \mathbb{R}^n$,
2. $u_t(x,t) + H(Du(x,t)) = 0$ *for a.e.* $(x,t) \in \mathbb{R}^n \times (0,\infty)$ *and*
3. $u(x+z,t) - 2u(x,t) + u(x-z,t) \leq C(1+\frac{1}{t})|z|^2$ *for some constant* $C \geq 0$ *and all* $x, z \in \mathbb{R}^n$, $t > 0$.

In particular, condition (3) will be vital in asserting uniqueness.

Theorem 10. *(Uniqueness of weak solutions). Assume H is C^2, convex and superlinear with g being Lipschitz. Then there exists at most one weak solution of the initial-value problem (2.4.4).*

We start by proving a preliminary result.

Lemma 8. *Let u be a weak solution (2.4.4) and I the $n \times n$ identity matrix. Then*

$$D^2 u^\epsilon(x,s) \leq C\left(1 + \frac{1}{s}\right) I,$$

for an appropriate constant $C \in \mathbb{R}$ and all $\epsilon > 0$, $x \in \mathbb{R}^n$ and $s > 2\epsilon$.

Proof. By semiconcavity of u, we have that for all $s > 0$ the mapping

$$x \to u(x,s) - \frac{C}{2}\left(1 + \frac{1}{s}\right)|x|^2$$

is concave. Thus $\frac{C}{2}\left(1+\frac{1}{s}\right)|x|^2 - u(x,s)$ is convex as is its mollification. We have:

$$\begin{aligned} 0 &\leq D^2\left(\left(\frac{C}{2}\left(1+\frac{1}{s}\right)|x|^2 - u(x,s)\right) * \eta_\epsilon\right) \\ &= D^2 \int_{B(0,\epsilon)} \left(\frac{C}{2}\left(1+\frac{1}{s}\right)|x-y|^2 - u(x-y,s-h)\right)\eta_\epsilon(y,h)\,dydh \\ &= D^2\left(\int_{B(0,\epsilon)} \left(\frac{C}{2}\left(1+\frac{1}{s}\right)|x-y|^2\eta_\epsilon(y,h)\,dydh\right.\right. \end{aligned}$$

$$
\begin{aligned}
&\quad - \int_{B(0,\epsilon)} u(x-y, s-h)\eta_\epsilon(y,h)\,dydh\Bigg) \\
&= C\left(1+\frac{1}{s}\right) D^2 \int_{B(0,\epsilon)} \frac{1}{2}|x-y|^2 \eta_\epsilon(y,h)\,dydh - D^2 u^\epsilon(x,s).
\end{aligned}
$$

Due to sufficient smoothness of our convex map, the derivative and convolution commute so that:

$$
\begin{aligned}
0 &\leq C\left(1+\frac{1}{s}\right) \int_{B(0,\epsilon)} D^2\left(\frac{1}{2}|x-y|^2\right)\eta_\epsilon(y,h)\,dydh - D^2 u^\epsilon(x,s) \\
&= C\left(1+\frac{1}{s}\right) \int_{B(0,\epsilon)} I\eta_\epsilon(y,h)\,dydh - D^2 u^\epsilon(x,s) \\
&= C\left(1+\frac{1}{s}\right) I - D^2 u^\epsilon(x,s).
\end{aligned}
$$

□

The proof of Theorem 10 is divided in to seven steps.

Proof. 1. Suppose that u and $\tilde{u}$ are weak solutions of (2.4.4) and define $w \coloneqq u - \tilde{u}$. Showing this equals zero almost everywhere proves the claim.

Observe that at any point (y,s) where both u and $\tilde{u}$ are differentiable and solve our PDE, we have

$$
\begin{aligned}
w_t(y,s) &= u_t(y,s) - \tilde{u}_t(y,s) \\
&= -H(Du(y,s)) + H(D\tilde{u}(y,s)) \\
&= -\int_0^1 \frac{d}{dr} H(rDu(y,s) + (1-r)D\tilde{u}(y,s))\,dr \\
&= -\int_0^1 DH(rDu(y,s) + (1-r)D\tilde{u}(y,s))\,dr \cdot (Du(y,s) - D\tilde{u}(y,s)) \\
&=: -b(y,s)\cdot Dw(y,s).
\end{aligned}
$$

Thus

$$
w_t + b\cdot Dw = 0 \quad \text{a.e.}
$$

2. Write $v \coloneqq \phi(w) \geq 0$, where $\phi : \mathbb{R} \to [0,\infty)$ is a smooth function to be selected later. Multiplying the above by $\phi'(w)$, we get

$$
v_t + b\cdot Dv = 0 \quad \text{a.e.}
$$

.

3. Choose $\epsilon > 0$ and define $u^\epsilon := \eta_\epsilon * u$, $\tilde{u}^\epsilon := \eta_\epsilon * \tilde{u}$, where η_ϵ is the standard mollifier in the x and t variables. According to the standard theory of mollifiers (see appendix C.4 from [1]), we have:

$$|Du^\epsilon| \leq \mathrm{Lip}(u), \quad |D\tilde{u}^\epsilon| \leq \mathrm{Lip}(\tilde{u})$$

with

$$Du^\epsilon \to Du, \quad D\tilde{u}^\epsilon \to D\tilde{u} \quad a.e., \text{ as } \epsilon \to 0.$$

Also by Lemma 8

$$D^2u^\epsilon, D^2\tilde{u}^\epsilon \leq C\left(1 + \frac{1}{s}\right)I$$

for an appropriate constant C and all $\epsilon > 0$, $y \in \mathbb{R}^n$, $s > 2\epsilon$.

4. Define

$$b_\epsilon(y,s) := \int_0^1 DH(rDu^\epsilon(y,s) + (1-r)D\tilde{u}^\epsilon(y,s))\,dr$$

so that

$$v_t + b_\epsilon \cdot Dv = (b_\epsilon - b) \cdot Dv \quad \text{a.e.}$$

and by the divergence product rule

$$v_t + \mathrm{div}(vb_\epsilon) = (\mathrm{div}\, b_\epsilon)v + (b_\epsilon - b) \cdot Dv \quad \text{a.e.}$$

5. Now since $H \in C^2(\mathbb{R}^n)$, we have

$$\begin{aligned}
\mathrm{div}\, b_\epsilon &= \sum_{l=1}^n \partial_l (b_\epsilon)_l = \sum_{l=1}^n \partial_l \int_0^1 H_l(rDu^\epsilon + (1-r)D\tilde{u}^\epsilon)\,dr \\
&= \int_0^1 \sum_{l=1}^n \partial_l (H_l(rDu^\epsilon + (1-r)D\tilde{u}^\epsilon))\,dr \\
&= \int_0^1 \sum_{l=1}^n (DH_l)(rDu^\epsilon + (1-r)D\tilde{u}^\epsilon) \cdot \partial_l (rDu^\epsilon + (1-r)D\tilde{u}^\epsilon)\,dr \\
&= \int_0^1 \sum_{l=1}^n \sum_{k=1}^n H_{kl}(rDu^\epsilon + (1-r)D\tilde{u}^\epsilon)(ru^\epsilon_{kl} + (1-r)\tilde{u}^\epsilon_{kl})\,dr \\
&\leq \int_0^1 \sum_{l=1}^n \sum_{k=1}^n C_0 C_1 \left(1 + \frac{1}{s}\right) \leq C\left(1 + \frac{1}{s}\right)
\end{aligned}$$

for some constant C in view of the inequalities of part 3 and convexity of H.

6. Fix $x_0 \in \mathbb{R}^n$, $t_0 > 0$, and set

$$R := \max\{|DH(p)| \,:\, |p| \le \max(\mathrm{Lip}(u), \mathrm{Lip}(\tilde{u}))\},$$

and define the cone

$$C := \{(x,t) \,:\, 0 \le t \le t_0,\ |x - x_0| \le R(t - t_0)\}.$$

Next write

$$e(t) = \int_{B(x_0, R(t_0 - t))} v(x,t)\, dx$$

so that

$$\begin{aligned}
\dot{e}(t) &= \int_{B(x_0,R(t_0-t))} v_t\, dx - R \int_{\partial B(x_0,R(t_0-t))} v\, dS \\
&= \int_{B(x_0,R(t_0-t))} -\mathrm{div}(vb_\epsilon) + \mathrm{div}(b_\epsilon)v + (b_\epsilon - b)\cdot Dv\, dx \\
&\quad - R \int_{\partial B(x_0,R(t_0-t))} v\, dS \\
&= -\int_{\partial B(x_0,R(t_0-t))} v(b_\epsilon \cdot \nu + R)\, dS \\
&\quad + \int_{B(x_0,R(t_0-t))} \mathrm{div}(b_\epsilon)v + (b_\epsilon - b)\cdot Dv\, dx \\
&\le \int_{B(x_0,R(t_0-t))} \mathrm{div}(b_\epsilon)v + (b_\epsilon - b)\cdot Dv\, dx \\
&\le C\left(1 + \frac{1}{t}\right) e(t) + \int_{B(x_0,R(t_0-t))} (b_\epsilon - b)\cdot Dv\, dx,
\end{aligned}$$

for a.e. $t > 0$. The first equality is the Leibniz integral rule, the second is the result from part 4 and the third is the divergence theorem. The first inequality is by parts 2, 3 and 4, and the last inequality follows from part 5. Let $\epsilon \to 0$ so that according to the inequalities and convergences of part 3, we can apply DCT above to get:

$$\dot{e}(t) \le C\left(1 + \frac{1}{t}\right) e(t) \quad \text{for a.e. } 0 < t < t_0.$$

7. Fix $0 < \epsilon < r < t$ and choose the function $\phi(x)$ to equal zero if

$$|z| \le \epsilon[\mathrm{Lip}(u) + \mathrm{Lip}(\tilde{u})]$$

and to be positive otherwise. Since $u = \tilde{u}$ on $\mathbb{R}^n \times \{t = 0\}$,

$$v(y, \epsilon) = \phi(w(y, \epsilon)) = \phi(u(y, \epsilon) - \tilde{u}(y, \epsilon)) = 0 \quad \text{at } \{t = \epsilon\},$$

because by the triangle inequality

$$\begin{aligned} |u(y, \epsilon) - \tilde{u}(y, \epsilon)| &\leq |u(y, \epsilon) - u(y, 0)| + |u(y, 0) - \tilde{u}(y, 0)| + |\tilde{u}(y, 0) - \tilde{u}(y, \epsilon)| \\ &= |u(y, \epsilon) - u(y, 0)| + |\tilde{u}(y, 0) - \tilde{u}(y, \epsilon)| \\ &\leq \epsilon(\text{Lip}(u) - \text{Lip}(\tilde{u})). \end{aligned}$$

Thus $e(\epsilon) = 0$ and by Grönwall's inequality (see appendix B.2 from [1]) we infer that

$$e(r) \leq e(\epsilon) \text{exp}\left(\int_{\epsilon}^{r} C\left(1 + \frac{1}{s}\right) ds \right) = 0.$$

Hence by the definition of ϕ we have:

$$|u(x, r) - \tilde{u}(x, r)| \leq \epsilon[\text{Lip}(u) + \text{Lip}(\tilde{u})] \quad \text{on } \ B(x_0, R(t_0 - r)),$$

for all $\epsilon > 0$. This means that $u \equiv \tilde{u}$ in $B(x_0, R(t_0 - r))$, so by continuity we have our claim $u(x_0, t_0) = \tilde{u}(x_0, t_0)$.

□

Combining the results of Lemma 5 and 6, and Theorem 10, we have our main result:

Theorem 11. *(Hopf-Lax formula as a weak solution). Suppose H is C^2, convex and superlinear with g being Lipschitz.. If either g is semiconcave or H is uniformly convex, then*

$$u(x, t) = \min_{y \in \mathbb{R}^n} \left\{ tL\left(\frac{x - y}{t}\right) + g(y) \right\}$$

is the unique weak solution of the initial-value problem (2.1.1) for the Hamilton-Jacobi equation.

Let us finish by having two examples of unique weak solutions using both the uniform convexity of H and the semiconcavity of the initial function g respectively.

Example 10. *Consider the initial-value problem:*

$$\begin{cases} u_t + \frac{1}{2}|Du|^2 = 0 & \text{in } \mathbb{R}^n \times (0, \infty) \\ u = |x| & \text{on } \mathbb{R}^n \times \{t = 0\}. \end{cases} \tag{2.4.5}$$

Here $H(p) = \frac{1}{2}|p|^2$ *so that* $L(v) = \frac{1}{2}|v|^2$. *The Hopf-Lax formula for the unique, weak solution of (2.4.5) is*

$$u(x,t) = \min_{y\in\mathbb{R}^n}\left\{\frac{|x-y|^2}{2t} + |y|\right\}.$$

Assume $|x| > t$. *Then for* $y \neq 0$

$$D_y\left(\frac{|x-y|^2}{2t} + |y|\right) = \frac{y-x}{t} + \frac{y}{|y|}.$$

This expression equals zero if $x = y + \frac{y}{|y|}t$, *and so* $y = (|x| - t)\frac{x}{|x|} \neq 0$. *Plugging this* y *in to the Hopf-Lax formula gives us:* $u(x,t) = |x| - \frac{t}{2}$ *if* $|x| > t$. *As for* $|x| \leq t$, *the minimum is attained at* $y = 0$. *Consequently*

$$u(x,t) = \begin{cases} |x| - \frac{t}{2} & \text{if } |x| \geq t \\ \frac{|x|^2}{2t} & \text{if } |x| \leq t. \end{cases}$$

The solution becomes semiconcave at times $t > 0$, *even though the initial function* $g(x) = |x|$ *is not semiconcave. This however is not an issue, since* H *being uniformly convex it is what Lemma 7 predicts.*

Example 11. *Consider the same problem with reversed initial conditions:*

$$\begin{cases} u_t + \frac{1}{2}|Du|^2 = 0 & \text{in } \mathbb{R}^n \times (0,\infty) \\ u = -|x| & \text{on } \mathbb{R}^n \times \{t = 0\}. \end{cases} \tag{2.4.6}$$

The Hopf-Lax formula for the unique, weak solution of (2.4.6) is

$$u(x,t) = \min_{y\in\mathbb{R}^n}\left\{\frac{|x-y|^2}{2t} - |y|\right\}.$$

Now for $y \neq 0$ *we have*

$$D_y\left(\frac{|x-y|^2}{2t} - |y|\right) = \frac{y-x}{t} - \frac{y}{|y|}.$$

This expression equals zero if $x = y - \frac{y}{|y|}t$, *and so* $y = (|x| + t)\frac{x}{|x|}$. *Plugging this* y *in to the Hopf-Lax formula gives us:*

$$u(x,t) = -|x| - \frac{t}{2}, \quad x \in \mathbb{R}^n, \ \ t \geq 0.$$

The initial function $g(x) = -|x|$ *is semiconcave, and the solution remains so for times* $t > 0$ *just as Lemma 6 predicts.*

Generalization and further discussion

In the more general case where H is neither convex nor lacking x-dependence, one can define other weak solutions to the Hamilton-Jacobi equations such as *viscosity solutions* (these actually apply to a much more general version called the Hamilton-Jacobi-Bellman equation). They are obtained from an approximate version of the Hamilton-Jacobi equation:

$$\begin{cases} u_t^\epsilon + H(Du^\epsilon, x) - \epsilon \Delta u^\epsilon &= 0 \quad \text{in } \mathbb{R}^n \times (0, \infty) \\ u^\epsilon &= g \quad \text{on } \mathbb{R}^n \times \{t = 0\}, \end{cases}$$

for $\epsilon > 0$. The point of this PDE is that whereas the original Hamilton-Jacobi equation (2.1.1) is fully non-linear, this one is quasilinear (and parabolic) and admits smooth solutions. This regularization effect is achieved via the extra term $\epsilon\Delta$. Then, roughly speaking, we let $\epsilon \to 0$ so that u^ϵ will converge to some weak solution of (2.1.1). This technique is known as the method of *vanishing viscosity*. The existence of such solutions u is a consequence of the optimality conditions of *control theory*, and surprisingly the Hopf-Lax formula (2.3.5) makes an appearance as a unique viscosity solution.

These topics can be found from Chapter 10 of [1], or more in depth as a whole in [4]. More on semiconcave functions, calculus of variations and the Legendre transform can be found in [3]. A classic reference to complete integrals and the method of characteristics is [5].

References

[1] Lawrence C. Evans. *Partial Differential Equations*, 2nd ed. American Mathematical Society, 2010. ISBN 978-0821849743.

[2] Hans Lewy. *Annals of Mathematics Second Series, Vol. 66, No. 1 (Jul., 1957), pp. 155-158 (4 pages).*

[3] P. Cannarsa and C. Sinestrani. *Semiconcave Functions, Hamilton-Jacobi Equations, and Optimal Control*, Volume 58. Birkhäuser, 2004. ISBN 0-8176-4336-2.

[4] M. Bardi and I. Capuzzo-Dolcetta. *Optimal Control and Viscosity Solutions of Hamilton-Jacobi-Bellman Equations*, Birkhäuser, 1997. ISBN 978-0-8176-4754-4.

[5] R. Courant and D. Hilbert. *Methods of Mathematical Physics, Vol. 2*, Wiley-Interscience, 1962. ISBN 978-0471504399.